

UTILISATION D'UNE COTATION D'ABONDANCE FONDÉE SUR UNE PROGRESSION GÉOMETRIQUE, POUR L'ANALYSE DES COMPOSANTES PRINCIPALES EN ÉCOLOGIE PLANCTONIQUE

SERGE FRONTIER

Centre O.R.S.T.O.M. de Nosy Be, Madagascar

et

FRÉDÉRIC IBANEZ

Station zoologique de Villefranche sur Mer, France

Résumé: Une analyse en composantes principales est effectuée d'une part sur un ensemble de variables biologiques (comptages d'organismes zooplanctoniques), d'autre part sur les valeurs de cotes d'abondance déduites de ces comptages. Les cotes correspondent à des intervalles de classes construits d'après une progression géométrique de raison 4.3. Les résultats des deux analyses sont extrêmement voisins, ce qui met en doute la nécessité de comptages précis en matière d'écologie du plancton.

Abstract: A principal component analysis is made upon a set of biological variates (counts of zooplankton organisms) and upon rough estimates of abundance of the same organisms given by means of a scale of classes. The classes are based upon a geometric progression with a basis of 4.3. The results of the two analyses are very similar. The authors conclude that the necessity of counting very precisely the plankton, in ecology, is doubtful.

Dans une note précédente Frontier (1969) avait suggéré de remplacer, dans les études d'écologie planctonique, les comptages exacts d'organismes par des estimations d'abondance fondées sur une progression géométrique de raison 4.3. Les classes d'abondance ainsi retenues étaient les suivantes:

Cotes d'abondance	Effectifs
1	1 à 3
2	4 à 17
3	18 à 80
4	80 à 350
5	350 à 1500
6	1500 à 6500
7	6500 à 27000
8	27000 à 120000
9	120000 à 500000
10	500000 à 2000000

Il est aisé de déterminer la classe d'abondance à laquelle appartient une catégorie d'organismes dans un échantillon, pratiquement sans erreur (la méthode pratique d'estimation importe peu). Il n'y a ambiguïté que si un effectif se trouve au voisinage d'une borne d'un des intervalles de classe; il convient alors d'adopter une cote décimale: par exemple '3.5' signifie 'environ 80'.

Quelques exemples concrets (Frontier, 1966, 1969, 1970) démontrent la coïncidence satisfaisante entre les résultats obtenus par comptage et par cotation d'abondance, dans la description de répartitions spatio-temporelles d'organismes planctoniques, à l'échelle d'observation généralement adoptée dans les études d'écologie régionale. On peut penser que pour des études plus particulières et plus précises, il conviendrait d'utiliser des comptages exacts ou approximatifs, ou encore des cotations d'abondance adaptées à chaque cas (c'est à dire établies en fonction de l'amplitude des variations significatives et de l'importance de la variabilité résiduelle).

L'analyse factorielle est considérée comme une technique fine. Toutefois, puisque le calcul traite globalement un nombre important de variables, chacune susceptible individuellement de donner lieu à une cotation d'abondance particulière, il a paru intéressant de vérifier si une cotation générale était susceptible de fournir des résultats voisins de ceux donnés par le traitement des comptages réels.

Dans le cadre de l'étude approfondie d'un secteur néritique tropical, une baie eutrophique de 800 km² (Baie d'Ambaro, côte nord-ouest de Madagascar) a été quadrillée par un réseau de 44 stations, en chacune desquelles était effectuée une récolte de plancton fond-surface. Une analyse en composantes principales a été réalisée à partir des résultats des comptages de 19 catégories zoologiques, et la même analyse a été exécutée sur l'ensemble des cotes d'abondance déduites des comptages. Les effectifs d'organismes (ramenés à 10 m³ d'eau filtrée) étaient préalablement transformés en leurs logarithmes en vue d'une normalisation approximative des variables. Les cotes d'abondance par contre, établies sur la base d'une progression géométrique et donc représentant des logarithmes, ont été traités sans transformation. La correction de Sheppard relative au groupement en classes n'a pas été appliquée.

Il a été ajouté une vingtième variable, calculée à partir d'une table de nombres au hasard. Il s'agit d'un test empirique de significativité des axes, dû à Ibanez (1973); cette variable, notée ε , est fortement exprimée par une des premières composantes principales, et l'on considère que seuls les axes extraits avant cette composante (et dans lesquels ε est très peu exprimée) ont une signification non fortuite.

Le résultat de la comparaison des deux analyses est résumé dans les tableaux et figures ci-après.

Le tableau I donne les éléments des quatre premiers vecteurs propres, normés à 1.0. On constate la coïncidence entre les vecteurs de même rang des deux analyses. La dernière variable est la variable-test ε : elle est fortement exprimée dans le quatrième vecteur, et très peu dans les précédents. On retiendra donc comme significatives les trois premières directions propres.

La coïncidence entre vecteurs propres de même rang dans les deux analyses peut

TABLEAU I

Les éléments des quatres premières vecteurs propres, normés à 1.0.

Espèces	Vecteurs propres (normés à 1)							
	1		2		3		4	
	A	B	A	B	A	B	A	B
1	-.01	.00	-.33	-.36	.04	.00	-.13	-.13
2	.31	.31	-.16	-.17	.00	-.04	.01	-.04
3	.11	.12	-.46	-.35	-.13	-.22	-.12	-.07
4	.25	.25	-.06	-.07	.04	.09	.05	.00
5	.19	.18	-.37	-.37	-.08	-.13	-.01	-.05
6	-.16	-.16	-.31	-.30	-.17	-.20	-.03	-.07
7	-.11	-.11	-.32	-.37	.18	.16	.38	.63
8	.01	.01	.13	.19	-.48	-.48	-.16	-.10
9	.00	.00	.13	.10	-.14	-.14	.60	.36
10	.02	.01	-.22	-.28	.51	.45	.14	.10
11	.33	.34	.05	.05	-.01	.02	.12	.12
12	.32	.32	-.16	-.15	-.17	-.16	.01	-.01
13	.31	.31	-.20	-.19	-.17	-.19	.00	-.01
14	-.04	-.02	.12	.20	-.47	-.42	.11	.05
15	.31	.30	-.07	-.04	-.14	-.16	-.03	-.07
16	-.32	-.33	-.15	-.12	-.01	-.05	-.06	-.06
17	-.32	-.32	-.17	-.16	-.07	-.09	-.08	-.09
18	-.25	-.25	-.22	-.21	-.22	-.28	.08	.00
19	-.28	-.29	-.21	-.18	-.17	-.17	-.04	-.05
20	-.05	-.03	-.08	-.09	-.16	-.18	.61	.62

être exprimée par la valeur de l'angle qu'ils forment dans un même espace R^{20} , la base étant l'ensemble des vecteurs 'espèces', orthonormée. Le cosinus de l'angle de deux vecteurs propres normés à 1 est le produit scalaire de ces vecteurs, immédiatement calculable. On trouve:

angle des vecteurs 1: 0° environ (cosinus trouvé un peu supérieur à 1 en raison de l'erreur d'arrondi)

angle des vecteurs 2: $9^\circ 30'$

angle des vecteurs 3: $7^\circ 55'$

angle des vecteurs 4: $9^\circ 50'$

En outre les valeurs propres sont très voisines dans les deux cas:

Valeurs propres	Comptages		Cotations	
	Variance	% de variance	Variance	% de variance
λ_1	7.63	38.17	7.51	37.57
λ_2	2.86	14.32	2.64	13.22
λ_3	1.90	9.50	1.87	9.35
λ_4	1.36	6.79	1.30	6.49

En conclusion, les deux nuages de points-observations dans R^{20} coïncident assez bien.

La contribution des différentes variables de départ à la constitution des vecteurs propres est généralement représentée par la position des vecteurs de la première base dans l'espace des saturations (espace rapporté aux vecteurs propres normés à leurs valeurs propres respectives). Nous représentons (Fig. 1) les projections des 20 vecteurs-espèces de R^{20} dans les plans des vecteurs propres 1 et 2, 1 et 3, 1 et 4, 2 et 3. Les points noirs sont obtenus à partir des comptages d'organismes, les points blancs à partir des cotations d'abondance. La proximité des deux analyses est manifeste.

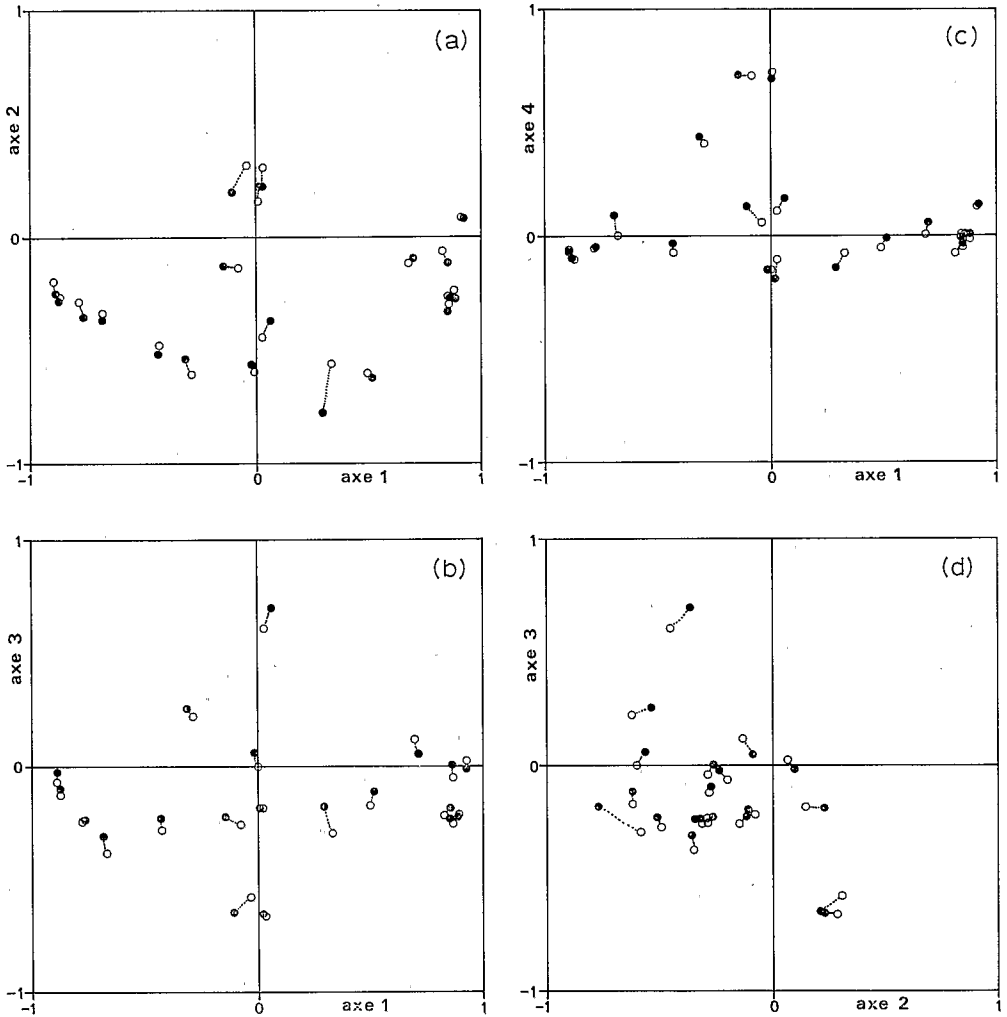


Fig. 1. Position des vecteurs-espèces dans l'espace des saturations. Points noirs, résultats obtenus à partir des comptages; points blancs, résultats obtenus à partir des cotations d'abondance.

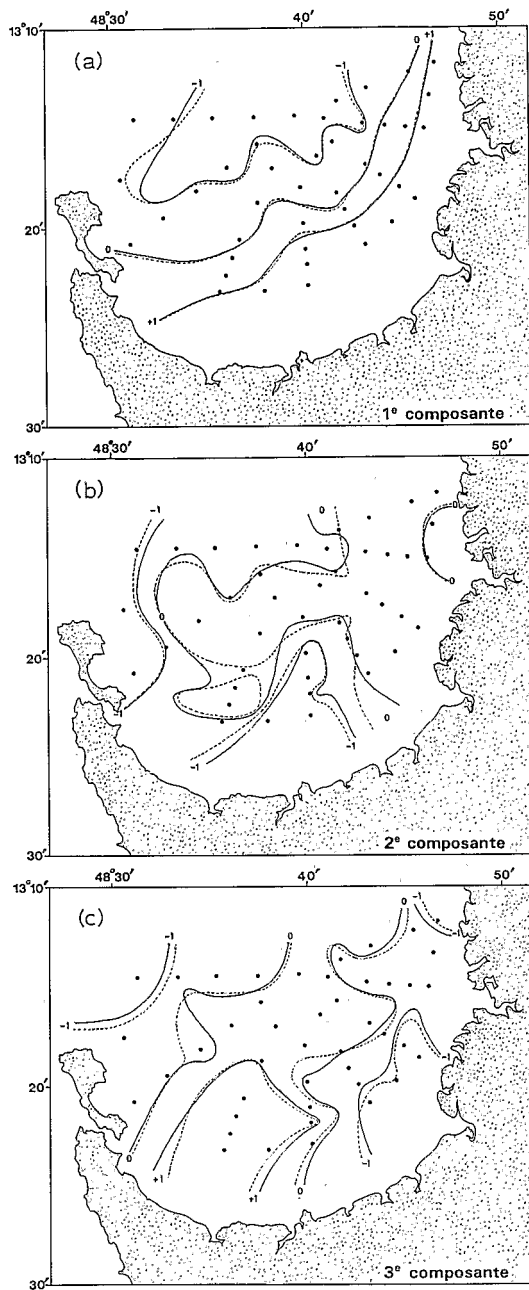


Fig. 2. Cartographies des trois premières composantes principales: trait plein, isolignes obtenues à partir des comptages; trait interrompu, isolignes obtenues à partir des cotations d'abondance.

Enfin, les valeurs prises par les différentes composantes principales aux 44 observations peuvent être portées sur la carte des stations. Il est généralement possible de tracer un certain nombre d'isolignes, ce qui révèle une certaine dépendance entre les 'facteurs principaux' et les coordonnées de l'espace physique. La Fig. 2 représente les isolignes -1 , 0 et $+1$ pour les trois premières composantes principales: celles obtenues à partir des comptages sont en trait plein, et celles obtenues à partir des cotations en trait interrompu. Ici encore la coincidence est très satisfaisante. Les coefficients de corrélation, calculés à partir des 44 observations, entre composantes principales de même rang dans les deux analyses, sont les suivants:

première composante: $r = 0.9977$

deuxième composante: $r = 0.9763$

troisième composante: $r = 0.9781$

quatrième composante: $r = 0.9838$

(angles de 4° , $12^\circ 30'$, 12° , $10^\circ 30'$ respectivement, dans \mathbf{R}^{44}).

Les différences constatées entre l'analyse des comptages et celle des cotations se révèle donc d'un ordre de grandeur inférieur à celui des fluctuations d'échantillonnage: aucune des différences notées entre les cartographies d'une composante ne saurait être confondue, à l'échelle d'observation choisie, avec une fluctuation à signification écologique.

Une difficulté apparaît dans le cas où deux des valeurs propres significatives sont très voisines. On sait que si deux valeurs propres sont égales il ne leur correspond pas de vecteurs propres mais un plan propre dans lequel il est possible de choisir arbitrairement deux vecteurs orthogonaux pour compléter la base des vecteurs propres. Si deux valeurs propres successives sont très voisines, l'incertitude de l'échantillonnage peut aboutir à une incertitude sur la direction des vecteurs propres correspondants. Dès lors l'erreur introduite en remplaçant les comptages exacts par des intervalles de classes peut se traduire par des changements de direction importants, et fortuits, des vecteurs propres de même rang (et même éventuellement à une inversion d'ordre)¹.

Ce cas s'est présenté lors d'un quadrillage de la baie réalisé à une date différente du précédent. L'analyse a donné les valeurs propres suivantes:

Valeurs propres	Comptages	Cotations
λ_1	7.81	7.67
λ_2	2.56	2.49
λ_3	2.35	2.28
λ_4	1.36	1.44

¹ Une légère différence de métrique (par exemple, transformation $\text{Arg Sh } \sqrt{x}$ au lieu de $\log(x+1)$) sur les données de départ, peut aboutir au même résultat.

Les angles des vecteurs propres de même rang, dans R^{20} , sont les suivants:

- vecteur 1: 0°
 vecteur 2: $22^\circ 40'$
 vecteur 3: $23^\circ 45'$
 vecteur 4: $14^\circ 15'$

En raison de la proximité des valeurs propres 2 et 3 les deux nuages de points dans R^{20} coïncident néanmoins approximativement. Mais la matrice des coefficients de corrélation entre les composantes des deux analyses calculés à partir des 44 observations met en évidence une certaine proportion d'échange entre les vecteurs 2 et 3:

		Composantes 1 ^{ère} analyse			
		1	2	3	4
	1	0.9981	<0.09	<0.09	<0.09
Composantes	2	<0.09	0.9144	0.3634	<0.09
2 ^{ième} analyse	3	<0.09	0.3563	0.9227	<0.09
	4	<0.09	<0.09	<0.09	0.9513

Dans ce cas particulier la cotation '4.3' n'aboutit pas en apparence aux mêmes résultats que les comptages réels. Cependant, il faut considérer que, dans l'analyse des comptages réels, 23.4 % de la variance totale se trouve liée au plan des vecteurs 2 et 3 sans que l'on puisse distinguer nettement ces deux derniers (dont l'orientation est, nous l'avons dit, fortement liée aux aléas de l'échantillonnage). La cartographie des composantes 2 et 3 prises séparément a donc peu de signification, et l'on doit considérer, dans l'analyse, la variance liée au plan dans son ensemble. Dans ces conditions les deux analyses coïncident.

En effet, on peut traiter globalement les vecteurs propres 2 et 3 en considérant dans R^{44} le vecteur à coordonnées complexes $W = V_2 + iV_3$, et son homologue W' pour la seconde analyse, les V_k étant les combinaisons linéaires des 20 espèces définies plus haut comme vecteurs propres normés à 1. On peut définir l'angle de deux vecteurs à coordonnées complexes par:

$$\cos \theta = \frac{\langle W, W' \rangle}{|W| \cdot |W'|}$$

On trouve immédiatement, en notant r_{jk} le coefficient de corrélation entre le $j^{\text{ième}}$ vecteur de la première analyse et de $k^{\text{ième}}$ vecteur de la seconde analyse:

$$|R| = |\cos \theta| = \frac{\sqrt{(r_{22} + r_{33})^2 + (r_{23} - r_{32})^2}}{2}$$

qui donne, dans notre exemple, la valeur 0.9865 (angle de $9^\circ 30'$)¹.

Cette étude a mis en évidence deux résultats fondamentaux: 1) d'une part la très grande stabilité de l'analyse des composantes principales; les directions indépendan-

¹ La phase du cosinus complexe dépend de la notation arbitraire du couple de vecteurs orthogonaux dans la plan propre.

tes qui ont un sens écologique sont invariables, même si l'on introduit une imprécision par une cotation d'abondance très large; et 2) d'autre part la possibilité de réaliser une économie de travail dans l'obtention d'une représentation factorielle des variables. Si l'analyse des composantes principales permet d'aboutir à une explication des phénomènes écologiques, elle peut servir également dans la phase exploratoire en donnant au chercheur des perspectives de travail. Il devient dès lors possible théoriquement et pratiquement, à l'aide d'un navire équipé d'un ordinateur de bord, d'analyser séquentiellement les communautés planctoniques au cours de la campagne.

BIBLIOGRAPHIE

- FRONTIER, S., 1966. Zooplancton de la région de Nosy Bé. I. Programme des récoltes et techniques d'étude. II. Plancton de surface aux stations 5 et 10. *Cah. O.R.S.T.O.M., sér. océanogr.*, T. 4, pp. 3-37.
- FRONTIER, S., 1969. Sur une méthode d'analyse faunistique rapide du zooplancton. *J. exp. mar. Biol. Ecol.*, Vol. 3, pp. 18-26.
- FRONTIER, S., 1970. Zooplancton de la région de Nosy Bé. IV. Plancton de surface aux stations 3, 4 et 11. *Cah. O.R.S.T.O.M., sér. océanogr.*, T. 8, pp. 3-17.
- IBANEZ, F., 1973. Méthode d'analyse spatio-temporelle du processus d'échantillonnage en planctologie; son influence dans l'interprétation des données par l'analyse en composantes principales. *Ann. Inst. océanogr. Paris, N.S.*, T. 43, pp. 83-111.